Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 2 (2010) 3586–3590

Procedia
Social and Behavioral Sciences

WCES-2010

Item transformation for computer assisted language testing: The adaptation of the Spanish University entrance examination

Jesus Garcia Laborda[a] *, Margarita Bakieva[a], José González-Such[a], Ana Sevilla Pavón[a]

[a]*Universidad de Alcalá, Department of Modern Philology, Alcalá de Henares, 28801, Spain*
[b] *U. Valencia Dpt.Mètodes d'Investigació i Diagnòstic en Educació, Blasco Ibáñez, 30, Valencia 46010, Spain*
[c]*Camille, Universitat Politecnica de Valencia, Camino de Vera, s/n, Valencia 46022, Spain*

Received October 28, 2009; revised December 4, 2009; accepted January 14, 2010

Abstract

Since the Spanish Educational system is changing and promoting the use of online tests, it was necessary to study the transformation of test items in the " Spanish University Entrance Examination " (IB P.A.U.) to diminish the effect of test delivery changes (through its computerization) in order to affect the least the current model. The purpose of this study was to describe and suggest the properties of a new test item taxonomy for the Spanish University Entrance Examination. After a convenient study and piloting by using previous research in computer-based language testing, the researchers created a taxonomy of test items for the I.B. P.A.U.

Keywords: Language test; item design; computers; educational change; washback.

1. Introduction

Since the Spanish Educational system is changing and promoting the use of online tests, it was necessary to study the transformation of test items in the " Spanish University Entrance Examination " (IB P.A.U.) to diminish the effect of test delivery changes (through its computerization) in order to affect the least the current model. The purpose of this study was to describe and suggest the properties of a new test item taxonomy for the Spanish University Entrance Examination. The study was fully funded by the Ministry of Education of Spain under the PAUER project (HUM2007-66479-C02-01/FILO).

Over the last twenty years the same paper model has been used to evaluate the foreign language competence in the University Entrance Examination (P.A.U.). The model based in old theories of unified competence (Oller, 1979; 1983) has experienced few changes in this time. Thus, the model is seen as old fashioned and its reputation has been decreasing especially in the last six or seven years. It is true that this criticism has been reflected in professional and research literature but no suggestions appeared until 2006 when the Universidad Politécnica de Valencia started two projects to design a computer based test. Certainly, this move also corresponds with similar ones across Europe, especially the Ösym test in Turkey. Leaded by similar incentives and final goals, both exams reflect with different

* Jesus Garcia Laborda. Tel.: +34-666-284557
E-mail address: jesusgalaborda@wanadoo.es.

1877-0428 © 2010 Published by Elsevier Ltd.
doi:10.1016/j.sbspro.2010.03.556

evolutions the need to increase the presence and, overall, the need to make the use of English (or, at least, foreign languages) present in daily life. It is in this manner that educational authorities see high stakes tests as a valid means to change the educational panorama and introduce educational innovation, especially in foreign languages (Wall, 2005). According to figure 1, there are two main perspectives from which educational change can be introduced: innovation oriented by the authorities and innovation claimed by the teachers. Of course, this is a simplistic perspective that leaves offside other stakeholders but this paper intends to show how these two aspects have the central role.

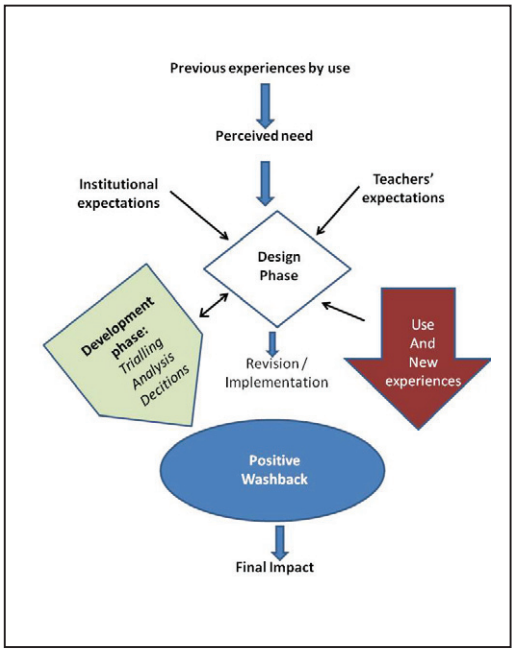

Figure 1: Designing new test items for P.A.U.

The current test of P.A.U. has been criticized by both teachers and administrators. In 2008 the Ministry of Education announced that a new format would be introduced in 2012. Little has been known from the Ministry about the future plans for the test especially considering its importance and the fact that the new test will introduce new oral tasks that may have a deep impact in the teaching methodology and syllabus design. Thus, although the current test (red arrow) is still in use and will bring new experiences to shape the new test tasks, teachers need to get ready for the new changes (of which so far in November 2009 they know nothing) while following the traditional preparation. At the same time (grey arrow) researchers are currently working towards new suggestions. Among these suggestions, lately the idea of a computer based test is gaining ground. This is how the Universidad Politécnica de Valencia (Spain) started a three-year research project to implement a new test. So far, most work has been done in the field of computer design and delivery. At this point, what seems really important is to obtain a series of tasks that could be included in the future computer based test.

2. Designing the new tasks

In their previous paper, Garcia Laborda & Gimeno Sanz (2008) planned to implement an adaptive testing system. In that case, most tasks should be graded similarly and have exactly the same style. It has been suggested in literature that since the students show a "Whole language competence" (Newman, 1985; Oller, 1979) just testing certain aspects of their competence would suffice to diagnose their foreign language competence. However, this trend has been considered wrong even by Oller himself (1983) and the latest advances in language testing have led to consider a completely different set of principles to design the new test items.

The first intention of the research team was to use the teachers' ideas to implement the new tasks. The idea was that since Garcia Laborda (2010a) detected a great deal of interest in changing the current paradigm and construct of the language test in a questionnaire given to 120 teachers from the city of Valencia and accounting for their

extended experience (most teachers were in their last 15 years before retirement), it was expected that their suggestions would bring new ideas into design phase. In order to facilitate the data selection, the questionnaire administered to 214 teachers in the Valencian Region (East of Spain) classified the tasks in four categories according to the foreign language use: speaking, listening, reading and writing. In order to design the method, a Delphi method was utilized with three main task designers as the original source of ideas and ten more teachers to provide feedback and propose changes. The items run from conservative tasks like multiple choice to innovative productive tasks abounding open ended questions. After the analysis, Gimeno Sanz et al. (2009) found that teachers preferred traditional tasks in three of the four skills but a very open question for the speaking skill. Overall, the computer based exam as suggested by the teachers consisted in the following sections:

a) Reading: Reading a text and answering three open questions or a set of multiple choice questions.

b) Writing: Writing a 200 word composition about a topic but little mention was given to doing more than one written piece with different registers or epistolary styles.

c) Listening: Listening to a two or three minute audio recording or video clip followed by a multiple choice section.

d) Students get a card and give a short (3 minute speech) on their own topic.

3. Moderating the teachers' attitudes and the communicative competence evaluation through the Delphi method

According to the test administrators in Valencia the results clearly evidence the teachers' reluctance to change the current format of the test supposedly due to a variety of reasons like fear to change, need to change teaching methodologies, fear of failure, incapability to approach the test from a different perspective or even some incapacity to cope with a more communicative approach. Since the goals of the educational reform under research are far more challenging and demanding with the current situation in the Spanish Educational system the administrators decided to submit the report, responses and proposals to a number of experts to moderate the responses and give recommendations for both further studies teachers' attitudes, in-school experimentation, teacher training and test construct design.

According to all of these premises the report was sent to two completely different teams of experts in language acquisition and educational measurement. The first team suggested including more open questions which may make the test more productive. They also acknowledged that communicative approaches are more difficult to achieve because these methodologies require full production and full understanding and testing strategies do not have such a significant role as in objective test types.

In order to achieve the moderating goal, the statistics team correlated the teachers' responses to the questionnaire to find that their answers were well linked. Given the responses of both teams their reactions and considerations are currently being under consideration. The Delphi method, however, seems to be an excellent means of providing a second view when extra opinions are necessary or just when moderating becomes the cornerstone of educational changes.

4. Results of moderation

As a result of the moderation process, the research team and the administrators observed that dramatic changes in language testing need to be gradually implemented in Spain. Currently, García Laborda et al. (in press) have observed that the best possibility to achieve a renovating situation is to combine both teachers' objective test items and more communicative ones. However, the idea would certainly not be by including both types of exercises but as García Laborda & Gimeno Sanz (2008) suggested, by mixing them in each exercise. This would also be the idea behind the new IB TOEFL format.

The new test for the University Entrance Examination that is currently being tested in Valencia includes the following tasks Appendix 1):

Skill	Task
Reading	1) Reading and check True/False response followed by a justification 2) Short communicative opinion based answers

	3)	Multiple choice questions
Writing		A 130-150 word composition based on personal attitudes and opinions
Listening	1)	Watching a mini-clip followed by open questions
	2)	Watching a mini-clip followed by a multiple choice exercise
Speaking	1)	Students watch a mini-clip and answer short questions
	2)	Students receive a prompt and then give a 2 minute mini-talk

Currently trialing is being carried with a simplified version in order to observe three main issues: ergonomics, washback in the language classroom and students' adaptation to the new environment.

5. Conclusions, findings and suggestions

The results provided a table and taxonomy of computer based items which resembles to those used in the TOEFL test but are clearly different from those used in IELTS. The experimentation also suggested possible sources of problems in the adaptation and singularities of computer based constraints. Current reactions from language teachers are also being collected by now with the hope that cooperation between the different stakeholders.

This study suggests that since educational change is extremely difficult, the Delphi method can be very valuable to design polls or taxonomies of items for computer based language testing, at least, it has been of great value in the case of Valencia.

References

García Laborda, J. & Gimeno Sanz, A. (2008) Adapatación del examen de inglés de las pruebas de acceso a la universidad a un entorno informático: Estudio sobre la tipología de preguntas. *Proceedings of the XXV congreso nacional de Lingüística Aplicada*, Murcia (Spain), XXV congreso nacional de lingüística aplicada, April 19-21: 19 de Abril al 21 de Abril de 2007: 723-730.

García Laborda, J. (2010) *Análisis de la formación del profesorado para la adaptación a una Prueba de Acceso a la Universidad informatizada*. Unpublished PhD dissertation. Universidad Complutense de Madrid, April 2010.

García Laborda, J.; Magal-Royo, T. & Enríquez Carrasco, E. (in press). Teachers' trialing procedures for Computer Assisted Language Testing Implementation. *Eurasian Journal of Educational Research*.

Gimeno Sanz, A.; Martínez Saéz, A. & Sevilla Pavón, A. (2009) *Resultados encuesta profesores 2º Bachillerato para la nueva prueba de lengua extranjera P.A.U.* Logse. Retrieved November 15, 2009 from http://www.upv.es/ingles/documentos/informe.pdf

Newman, J. M. (de.) (1985). *Whole Language Theory in Use*. London: Heineman.

Oller, J. W. (1979). *Language Tests at School*. A Pragmatic Approach. Harlow, Essex: Longman.

Oller, J. W. (1983). Issues in Language Testing Research. Rowley, Mass.: Newbury House.

Wall, D. (2005). *Studies in language testing 22: The impact of high-stakes testing on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge: Cambridge University

Appendix 1

Table 1: Descriptive statistics of the percentages obtained by items and skills

	N		Average	Median	Mode	Standard Deviation	Variance	Minimum	Maximum	Total
	Valid	Lost/Null	Valid	Lost/Null	Valid	Lost/Null	Valid	Lost/Null	Valid	Lost/Null
	214	0	2,19	3,00	3	,907	,822	1	3	469
Reading A	214	0	6,60	8,00	2	3,491	12,185	0	12	1412
Reading B	214	0	8,27	8,00	10	3,006	9,034	0	12	1770
Reading C	214	0	8,26	10,00	12	3,553	12,626	0	12	1768
Reading D	214	0	6,77	6,00	4(a)	3,397	11,541	0	12	1448
Reading E	214	0	5,68	5,00	4	3,174	10,077	0	12	1216
Reading F	214	0	5,51	6,00	4	3,013	9,077	0	12	1180
Writing A	214	0	8,21	10,00	10	3,166	10,023	0	12	1756
Writing B	214	0	9,49	12,00	12	3,453	11,922	0	12	2030
Writing C	214	0	7,06	8,00	8	3,044	9,264	0	12	1510
Writing D	214	0	5,47	6,00	6	2,945	8,673	0	12	1170
Writing E	214	0	5,28	4,00	4	2,779	7,724	0	12	1130
Writing F	214	0	5,68	6,00	4	3,174	10,077	0	12	1216
Listening A	214	0	6,70	6,00	2	3,549	12,596	0	12	1434
Listening B	214	0	8,33	10,00	10	3,194	10,202	0	12	1782
Listening C	214	0	8,88	10,00	8	2,695	7,262	0	12	1900
Listening D	214	0	5,69	4,00	2	3,687	13,595	0	12	1218
Listening E	214	0	5,89	6,00	4	2,885	8,325	0	12	1260
Listening F	214	0	5,55	6,00	2	3,043	9,263	0	12	1188
Speaking A	214	0	5,35	4,00	2	3,795	14,405	0	12	1144
Speaking B	214	0	7,67	8,00	10	3,165	10,015	0	12	1642
Speaking C	214	0	7,86	8,00	12	3,702	13,708	0	12	1682
Speaking D	214	0	7,45	8,00	8	2,896	8,389	0	12	1594
Speaking E	214	0	6,84	6,00	4	3,483	12,134	0	12	1464
Speaking F	214	0	5,08	4,00	2	3,188	10,162	0	12	1088